JOURNAL OF VIROLOGY, May 2008, p. 5099–5103 Vol. 82, No. 10
0022-538X/08/$08.00+0 doi:10.1128/JVI.01993-07

Genomic Analysis of Closely Related Astroviruses▽

Errol Strain,[1] Laura A. Kelley,[3] Stacey Schultz-Cherry,[4] Spencer V. Muse,[1] and Matthew D. Koci[2]*

Department of Statistics[1] and Department of Poultry Science,[2] North Carolina State University, Raleigh, North Carolina 27695; Plant Science Research, Biosciences Research Laboratory, USDA-ARS, Fargo, North Dakota 58105[3]; and Department of Medical Microbiology and Immunology, University of Wisconsin, Madison, Wisconsin 53706[4]

Received 10 September 2007/Accepted 23 February 2008

To understand astrovirus biology, it is essential to understand factors associated with its evolution. The current study reports the genomic sequences of nine novel turkey astrovirus (TAstV) type 2-like clinical isolates. This represents, to our knowledge, the largest genomic-length data set available for any one astrovirus type. The comparison of these TAstV sequences suggests that the TAstV species contains multiple subtypes and that recombination events have occurred across the astrovirus genome. In addition, the analysis of the capsid gene demonstrated evidence for both site-specific positive selection and purifying selection.

Members of the *Astroviridae* are frequently associated with clinical diarrhea in the young of both mammalian and avian hosts (2, 9, 15). In spite of their worldwide distribution and endemic nature, our understanding of their evolution is limited. The majority of previous studies of astrovirus phylogeny have examined relationships among clinical isolates based on diagnostic reverse transcription-PCR amplicons (5, 19). While sequences from short diagnostic amplicons have been successful in assigning isolates to groups, they lack the power to accurately resolve phylogenetic relationships (14). Studies which have attempted to reconstruct the astrovirus phylogeny using genome alignments have done so across astrovirus genera (26). The sequence divergence and differences in codon usage across the *Astroviridae* may confound conclusions about phylogenetic relationships and selection pressure. The current study describes the phylogenetic analysis of multiple genomic sequences of closely related turkey astrovirus (TAstV) clinical isolates collected from commercial turkey flocks across the United States. Representative isolates from each location were randomly selected, and the full-length genomic sequences were determined as previously described (8). Chromatogram data were analyzed using phred, phrap, and consed software (3, 4); sequences were aligned using ClustalW (25) and were edited using GeneDoc (18). Amino acid- and nucleotide-based estimates of phylogeny were generated by using both MrBayes (6) and the hypothesis testing using phylogenies (HYPHY) package (10).

Evidence of distinct subtypes. The relationships of the novel clinical TAstV isolates (GenBank accession numbers EU143843 to EU143851) within the *Astroviridae* were first assessed by using predicted capsid amino acid sequences. The topology of this tree (Fig. 1A) was consistent with previous studies demonstrating two major clades containing the genera *Mamastrovirus* and *Avastrovirus* and minor clades corresponding to their host species (29). All of the clinical isolates clustered with TAstV-2/NC/99, TAstV1987, and TAstV2001 (TAstV-2-like) (Fig. 1A). The TAstV-1 capsid sequence was found in a clade with avian nephritis virus, with the distance between the reference TAstV-1 (accession no. CAB95007) and TAstV-2 (TAstV-2/NC/99; accession no. AAF18464) sequences comparable to the distance between human astroviruses (HAstVs) and other mamastroviruses. The sequence analysis of TAstV-1 and TAstV-2 diagnostic amplicons, previously described by Pantin-Jackwood et al. (19) and Cattoli et al. (1), demonstrated that the levels of variation among TAstV-1-like isolates and among TAstV-2-like isolates are comparable to the level of diversity among HAstVs. The phylogenetic analysis of the full-length capsid genes of all TAstV viruses (Fig. 1A) suggests that TAstV-1-like and TAstV-2-like viruses may have originated from separate introductions into the turkey species and that there are at least two TAstV lineages which should be regarded as distinct subtypes instead of serotypes. Within each subtype, there appears the potential for distinct serotypes to exist, as TAstV2001 and TAstV1987 have been reported to represent distinct serotypes (24) and share only 73% nucleotide sequence identity (23). This level of sequence conservation is similar to that of HAstV capsid genes from different serotypes (<80% nucleotide similarity; unpublished observation). These sequence differences suggest that MN/01 may represent a serotype that is distinct from that of TAstV-2/NC/99; however, experimental examination of the serological cross-reactivity of MN/01 with other viruses is needed. Collectively, these findings suggest that the ecology of *Avastrovirus* species may be more complicated than currently appreciated. Interestingly, Lukashov et al. (14) described the phylogenetic evidence of at least two cross-species transmissions within the genus *Mamastrovirus*. This leads one to question if other, as-yet-unidentified astrovirus subtypes exist within mammalian populations.

Genomic analysis is required to understand phylogeny. To develop a more-accurate reconstruction of the relationships among the TAstV-2-like viruses, phylogenies were constructed using genomic, open reading frame 1a (ORF1a), ORF1b, and ORF2 sequences from the TAstV-2-like clinical isolates and

* Corresponding author. Mailing address: Department of Poultry Science, North Carolina State University, Campus Box 7608, Raleigh, NC 27695. Phone: (919) 515-5388. Fax: (919) 515-2625. E-mail: mdkoci@ncsu.edu.

▽ Published ahead of print on 5 March 2008.

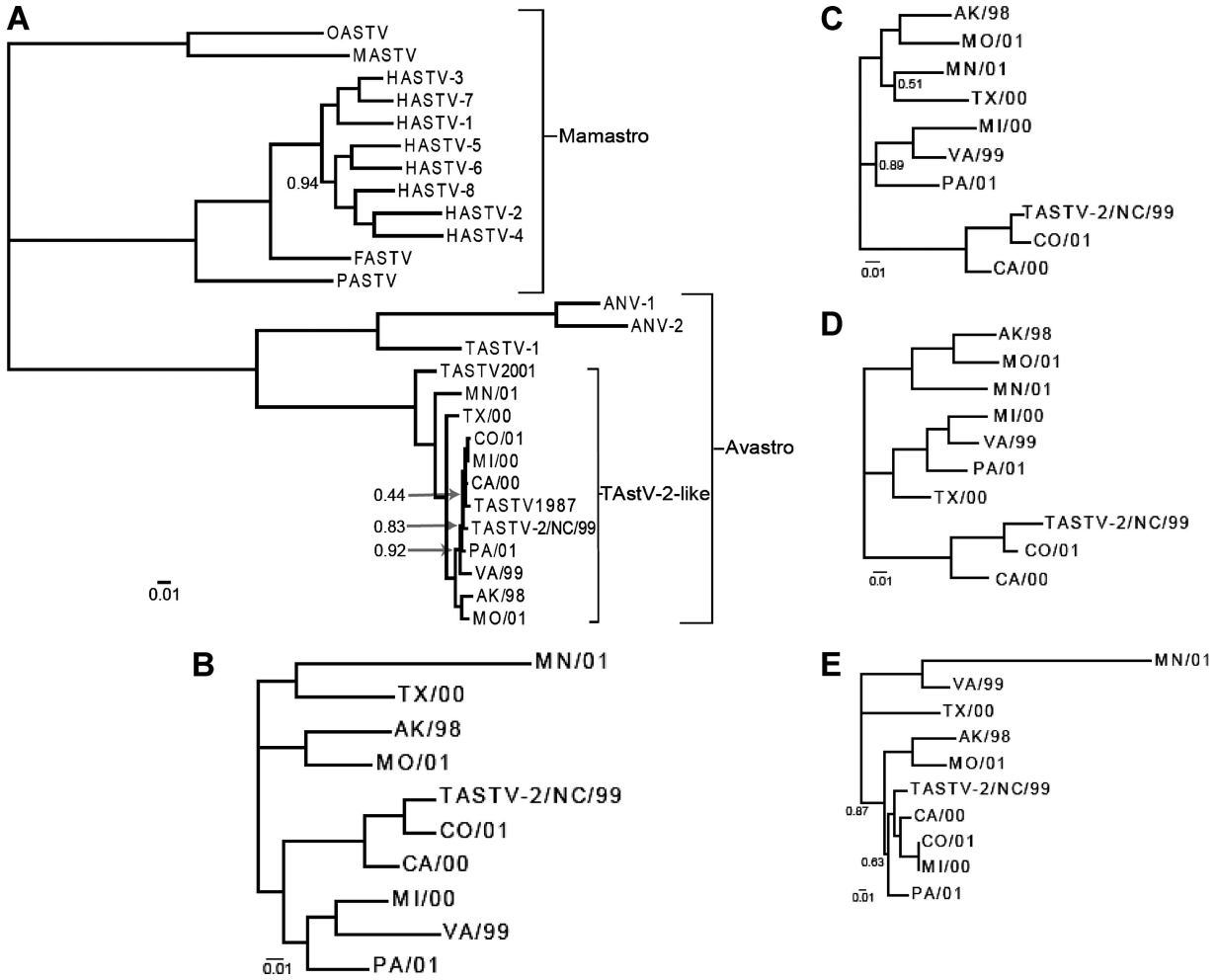

FIG. 1. Phylogenetic relationship of TAstV-2-like viruses. (A) A Bayesian phylogenetic tree describing evolutionary relationships among predicted amino acid sequences for human (HAstV), pig (PAstV), sheep (OAstV), mink (MAstV), avian (ANV), and turkey (TAstV) astrovirus capsids. (B to E) Bayesian phylogenetic trees describing the nucleotide relationships among the newly reported TAstV-2-like isolates and the TAstV-2/NC/99 reference sequence. Alignments were constructed using whole genomes (B), ORF1a (C), ORF1b (D), and ORF2 (E). Branch lengths represent the expected number of amino acid substitutions (A) and the expected number of nucleotide substitutions per site (B to E). Posterior support values are shown only for nodes with values less than 0.95. Estimates of phylogeny were made by using the MrBayes program (6).

the TAstV-2/NC/99 reference sequence (Fig. 1B to E). The analysis of the TAstV-2-like clinical isolates demonstrates variation in phylogenetic relationships across the different ORFs in comparison to the full genome (Fig. 1B to E). While generating this level of data on a routine basis is impractical for diagnostic purposes, it is important to recognize that the region of viral genome analyzed can affect the interpretation of phylogenetic relationships. The initial characterization of a virus based on its capsid sequence is effective for establishing its genus and species; however, to understand the evolutionary history of an isolate during an outbreak, sufficient sequence coverage should be included to ensure the most-accurate relationship possible. This is highlighted by the observation that MI/00 clustered with PA/01 and VA/99 in ORF1a and ORF1b trees (Fig. 1C to D) but was found with CO/01 in the capsid phylogeny tree (Fig. 1E). MI/00 and CO/01 ORF2 sequences were 99% identical for both amino acids and nucleotides, while the ORF1a and -1b sequences had ≥91% nucleotide and ≥94% amino acid identities, respectively. These observations, together with reports by Walter et al. (28) and Pantin-Jackwood et al. (20), suggest that the region around the ORF1b-ORF2 junction is a potential recombination hot spot.

Phylogenetic evidence of recombination across the astrovirus genome. The Sawyer test for recombination (21) was performed to further analyze the potential recombination event between MI/00 and CO/01, and a breakpoint was identified ($P < 0.0001$) at nucleotide position 4861. To determine if this region was the only region associated with recombination, the analysis was expanded using GENECONV (21) to test all pairwise comparisons of the entire isolate genomes. Forty-six total recombination events were identified, with at least one recombination event identified in each of the 10 TAstV-2-like isolates (accession numbers EU143843 to EU143851 and AF206663) (Fig. 2A). The distribution of the putative recombination events corresponded with the level of divergence across the three reading frames. ORF1b is the least divergent and had only two putative recombination events. ORF2 is the most divergent and contained the majority of putative recom-

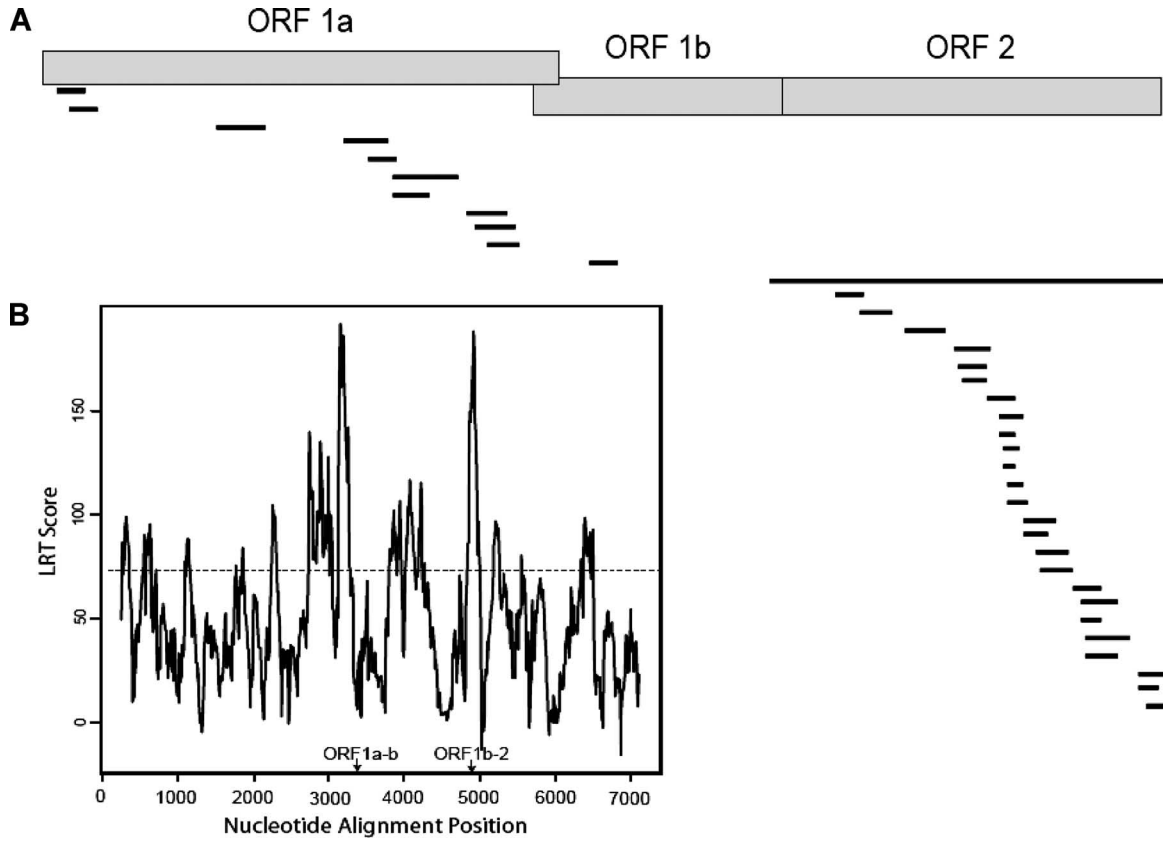

FIG. 2. Phylogenetic evidence for recombination across TAstV-2-like virus genomes. (A) All pairwise comparisons of the TAstV-2-like genomic sequences were tested by using GENECONV to identify the sites of recombination between isolates ($P < 0.05$). Gray rectangles represent the TAstV-2 genome and individual ORFs. The start and end of each black line correspond to the relative locations within the genome of the start and end of the recombination breakpoint as identified by GENECONV. (B) The analysis of recombination in the multiple alignments of TAstV-2-like isolates was also performed by using the TOPALi program. On the *x* axis, the nucleotide position within the genome is shown, while the likelihood ratio (LRT) score indicating support for recombination is shown on the *y* axis. The LRT score corresponds to the difference in likelihood between models that restrict recombination and models that allow for recombination at a particular nucleotide position. LRT scores above the dashed horizontal line are significant at a *P* of 0.05.

bination events (Fig. 2A). The evidence of recombination was also assessed using TOPALi (16) to analyze the TAstV-2-like multiple alignment. TOPALi analysis demonstrated similar evidence for recombination, with the strongest support near the junctions between ORFs (Fig. 2B), suggesting an association between recombination and transcriptional signal sequences (7, 13). Furthermore, the finding that at least one putative recombination event was detected in every isolate suggests that recombination may play a key role in astrovirus sequence diversity.

Role of selection in astrovirus capsid evolution. Sequence diversity in astroviruses may also involve host selection pressures. These pressures would presumably explain the existence of distinct serotypes. To address this, TAstV-2-like astroviruses were analyzed for selection using two alignments. The first alignment contained TAstV-2-like capsid sequences (accession numbers EU143843 to EU143851 and AF206663). The second alignment contained TAstV-2-like sequences that were found to have >80% nucleotide similarity to TAstV-2/NC/99, excluding sequences which may belong to different serotypes (MN/01) based on the sequence distances observed (Fig. 1A). For comparison, two HAstV alignments were also analyzed. The first included at least one sequence from each serotype (HAstV), and the second alignment included eight HAstV-4 capsid sequences (HAstV-4). The phylogenies for the four alignments were constructed using HYPHY and analyzed by genetic algorithm for recombination detection (GARD) (11). Breakpoints were recorded in the data input file, and tests for selection were performed using the fixed-effects likelihood (FEL), internal fixed-effects likelihood (IFEL), and random-effects likelihood (REL) models (10, 11) and the partitioning approach for robust inference of selection (PARRIS) (22) methods implemented at http://www.datamonkey.org. Each method uses likelihood-based analysis to identify sites where the rate of nonsynonymous substitution is greater than the rate of synonymous substitution.

The FEL and IFEL methods identified a small number of positively selected sites in the TAstV-2 alignments, while REL and PARRIS found no evidence of positive selection (Tables 1 and 2). More sites were identified as positively selected with IFEL than FEL, suggesting that selective pressure is occurring primarily at the population level (internal branches). Results from the analysis of the HAstV were similar to those for TAstV-2, except that REL identified positively selected sites in the multiple-serotype HAstV data set (Table 1). Sites with

TABLE 1. Sites (codons) selected by different methods

Isolate[a]	No. of aa aligned[b]	No. of sites[c]							
		Positive selection				Purifying selection			
		FEL	IFEL	REL	PARRIS	FEL	IFEL	REL	PARRIS
HAstV	727	5 (1)	6 (3)	29 (2)	0	496 (433)	479 (416)	545 (450)	ND[e]
HAstV-4	727	1 (0)	1 (0)	0 (0)	0	138 (42)	65 (35)	727[d]	ND
TastV-2-like	771	1 (0)	9 (5)	0 (0)	0	236 (130)	107 (49)	771[d]	ND
TAstV-2	804	4 (1)	6 (3)	0 (0)	0	178 (86)	56 (25)	804[d]	ND

[a] Alignments were done for capsid sequences from each serotype for HAstV, capsid sequences for HAstV-4, capsid sequences in the TAstV-2-like clade (TastV-2-like) (Fig. 1), and TastV-2-like capsid sequences with >80% similarity to TastV-2/NC/99 (TAstV-2).
[b] aa, amino acids.
[c] Number of significant sites at a P of ≤0.1 ($P \leq 0.05$), except where indicated otherwise.
[d] Number of significant sites at a P of ≤0.05.
[e] ND, not determined.

strong evidence for selection in TAstV-2 and HAstV were located primarily near the 3′ end of the sequence, which is thought to comprise the outer surface of the viral capsid (12).

FEL, IFEL, and REL were also used to look for sites that were under purifying selection, where the rate of synonymous substitution is higher than the rate of nonsynonymous substitution. Considerably more sites were identified to be under purifying selection (Table 1), similar to previous findings (27). The interpretation of the selection analysis may be confounded by the fact that the models of sequence evolution used in these analyses assume that the alignment is fixed and may not detect positive selection at indel sites. This may explain why the more-diverse data sets show more sites under purifying selection. Selection pressures (selective pressures conferring an advantage to mutant viruses) are likely to be involved in astrovirus capsid evolution; however, these data (Tables 1 and 2) suggest that positive selection at the codon level is not the dominant mechanism driving diversity.

This study is the first, to our knowledge, to analyze the phylogenetic relationships of multiple, highly related, full-length astrovirus genomes and the first to explore the role of recombination and selection across the genomes of highly related astroviruses. The results from these analyses suggest there are two distinct subtypes of TAstV which may have resulted from separate introductions into turkeys. Overall, the results from these analyses do not point to one mechanism as the primary means of achieving sequence diversity in astroviruses; instead, they suggest that astroviruses employ all sequence-changing mechanisms available to positive-sense single-stranded RNA viruses and underscore the need for models which allow for all of these factors to be analyzed together.

TABLE 2. Positively selected sites in TAstV-2 isolates

Codon position in TAstV-2 sequence	P value for positive selection[a]			
	TAstV-2-like		TAstV-2 (>80%)[b]	
	FEL	IFEL	FEL	IFEL
12		0.080		
61		0.074		
365		0.014		
387		0.010	0.098	
391		0.076		0.082
473				0.081
477		0.083		
485			0.091	
493				0.048
598	0.060	0.015	0.085	0.023
602		0.043		0.072
636		0.033	0.042	0.025

[a] Only P values of ≤0.1 are shown.
[b] Sequences were found to have a >80% similarity to TAstV-2/NC/99.

Accession numbers. Accession numbers for capsid amino acid sequences are as follows: CAB95007 (TAstV-1), AAV37187 (TAstV2001), AAV37186 (TAstV1987), AAF18464 (TAstV-2/NC/99), BAA92849 (ANV1), BAB21617 (ANV-2), NP_059946 (ovine AstV), NP_795336 (mink AstV), CAB95000 (porcine AstV), and AAC13556 (feline AstV). Accession numbers for nucleotide sequences are as follows: AF206663 (TAstV-2/NC/99), AY769615 (TAstV1987), AY769616 (TAstV2001), AY720892 (HAstV-1), AY720892 (HAstV-1 ORF2), L23513 (HAstV-1 ORF2), L13745 (HAstV2), L06802 (HAstV-2 ORF2), AF292074 (HAstV-3 ORF1b), AF117209 (HAstV-3 ORF2), AY720891 (HAstV-4), AB025801 (HAstV-4 ORF2), AB025802 (HAstV-4 ORF2), AB025804 (HAstV-4 ORF2), DQ344027 (HAstV-4 ORF2), DQ070852 (HAstV-4 ORF2), Z33883 (HAstV-4 ORF2), DQ028633 (HAstV-5), AB037273 (HAstV-5 ORF2), AB037274 (HAstV-5 ORF2), U15136 (HAstV-5 ORF2), AF292077 (HAstV-6 ORF1b), Z46658 (HAstV-6 ORF2), AF248738 (HAstV-7 ORF1b), Y08632 (HAstV-7 ORF2), AF248738 (HAstV-7 ORF2), AF260508 (HAstV-8), Z66541 (HAstV-8 ORF2), AB031031 (HAstV Katano23-6), AB031030 (HAstV Katano24), AF141381 (HAstV ORF2, unclassified serotype), and AB013618 (HAstV ORF2, unclassified serotype). Accession numbers for novel isolates are as follows: EU143843 (TAstV/AK/98), EU143844 (TAstV/CA/00), EU143845 (TAstV/CO/01), EU143846 (TAstV/MI/01), EU143847 (TAstV/MN/01), EU143848 (TAstV/MO/01), EU143849 (TAstV/PA/01), EU143850 (TAstV/TX/00), and EU143851 (TAstV/VA/99).

We thank Chris Ashwell, Shelly Nolin, and Audrey O'Nan of the Domestic Animal Genomics Laboratory at North Carolina State University for their help with sequencing and Sergei L. Kosakovsky Pond, Barb Sherry, and Liz Turpin for their many helpful discussions and critical reviews of the manuscript.

This work was supported in part by P30 DK34987 and the North Carolina Agricultural Foundation.

REFERENCES

1. **Cattoli, G., C. De Battisti, A. Toffan, A. Salviato, A. Lavazza, M. Cerioli, and I. Capua.** 2007. Co-circulation of distinct genetic lineages of astroviruses in turkeys and guinea fowl. Arch. Virol. **152:**595–602.
2. **Caul, E. O., and H. Appleton.** 1982. The electron microscopical and physical

characteristics of small round human fecal viruses: an interim scheme for classification. J. Med. Virol. **9:**257–265.
3. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
4. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
5. **Guix, S., A. Bosch, and R. M. Pinto.** 2005. Human astrovirus diagnosis and typing: current and future prospects. Lett. Appl. Microbiol. **41:**103–105.
6. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
7. **Jonassen, C. M., T. T. Jonassen, T. M. Sveen, and B. Grinde.** 2003. Complete genomic sequences of astroviruses from sheep and turkey: comparison with related viruses. Virus Res. **91:**195–201.
8. **Koci, M. D., B. S. Seal, and S. Schultz-Cherry.** 2000. Molecular characterization of an avian astrovirus. J. Virol. **74:**6173–6177.
9. **Koopmans, M. P., M. H. Bijen, S. S. Monroe, and J. Vinje.** 1998. Age-stratified seroprevalence of neutralizing antibodies to astrovirus types 1 to 7 in humans in The Netherlands. Clin. Diagn. Lab. Immunol. **5:**33–37.
10. **Kosakovsky Pond, S. L., and S. D. Frost.** 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. **22:**1208–1222.
11. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23:**1891–1901.
12. **Krishna, N. K.** 2005. Identification of structural domains involved in astrovirus capsid biology. Viral Immunol. **18:**17–26.
13. **Lewis, T. L., and S. M. Matsui.** 1997. Studies of the astrovirus signal that induces (−1) ribosomal frameshifting. Adv. Exp. Med. Biol. **412:**323–330.
14. **Lukashov, V. V., and J. Goudsmit.** 2002. Evolutionary relationships among *Astroviridae*. J. Gen. Virol. **83:**1397–1405.
15. **McNulty, M. S., W. L. Curran, and J. B. McFerran.** 1980. Detection of astroviruses in turkey faeces by direct electron microscopy. Vet. Rec. **106:**561.
16. **Milne, I., F. Wright, G. Rowe, D. F. Marshall, D. Husmeier, and G. McGuire.** 2004. TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. Bioinformatics **20:**1806–1807.
17. Reference deleted.
18. **Nicholas, K. B., H. B. J. Nicholas, and D. W. I. Deerfield.** 1997. GeneDoc: analysis and visualization of genetic variation. EMBNet News **4:**14.
19. **Pantin-Jackwood, M. J., E. Spackman, and P. R. Woolcock.** 2006. Molecular characterization and typing of chicken and turkey astroviruses circulating in the United States: implications for diagnostics. Avian Dis. **50:**397–404.
20. **Pantin-Jackwood, M. J., E. Spackman, and P. R. Woolcock.** 2006. Phylogenetic analysis of turkey astroviruses reveals evidence of recombination. Virus Genes **32:**187–192.
21. **Sawyer, S.** 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:**526–538.
22. **Scheffler, K., D. P. Martin, and C. Seoighe.** 2006. Robust inference of positive selection from recombining coding sequences. Bioinformatics **22:**2493–2499.
23. **Tang, Y., A. M. Murgia, and Y. M. Saif.** 2005. Molecular characterization of the capsid gene of two serotypes of turkey astroviruses. Avian Dis. **49:**514–519.
24. **Tang, Y., and Y. M. Saif.** 2004. Antigenicity of two turkey astrovirus isolates. Avian Dis. **48:**896–901.
25. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
26. **van Hemert, F. J., B. Berkhout, and V. V. Lukashov.** 2007. Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the *Astroviridae*. Virology **361:**447–454.
27. **van Hemert, F. J., V. V. Lukashov, and B. Berkhout.** 2007. Different rates of (non-)synonymous mutations in astrovirus genes; correlation with gene function. Virol. J. **4:**25.
28. **Walter, J. E., J. Briggs, M. L. Guerrero, D. O. Matson, L. K. Pickering, G. Ruiz-Palacios, T. Berke, and D. K. Mitchell.** 2001. Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children. Arch. Virol. **146:**2357–2367.
29. **Walter, J. E., and D. K. Mitchell.** 2003. Astrovirus infection in children. Curr. Opin. Infect. Dis. **16:**247–253.